Model selection methods

Goodness-of-fit criteria:

① $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$

② $R^2_{adj} = 1 - \frac{MSE}{MST} = 1 - \frac{n-1}{n-p}(1-R^2)$

③ AIC = Akaike Information Criterion

$= n \ln\left(\frac{SSE}{n}\right) + 2p$ [smaller is better]

④ Mallow's $C_p = \frac{SSE_p}{MSE_{full}} - (n-2p)$

If the model with $p = k+1$ parameters is approximately as good as the full model, then

$MSE_p \approx MSE_{full}$

$\parallel$

$\frac{SSE_p}{n-p}$

$\Rightarrow C_p \approx \frac{(n-p)\,MSE_{full}}{MSE_{full}} - (n-2p) = p$

Stepwise methods

① Forward selection
- Evaluate each 1-predictor model & select the best one
- Keep that predictor & evaluate each 2-predictor model, selecting the best one
- replat until stopping point (discuss later)

If you fit all possible models, you would have to run $2^k - 1$ models

Forward selection would require running

$$k + (k-1) + (k-2) + \dots + 1 = \frac{k(k+1)}{2}$$

② Backward elimination
- Fit the model with all possible predictors
- Fit every possible model with 1 predictor omitted & select the best one
- Fit every possible model that omits the first predictor that was chosen as well as a second predictor
- repeat until stopping point

(3) Stepwise method

- Go 2 steps by forward selection
- Perform a backward step
- Perform a forward step
- Perform a backward step
- Continue until stopping point

What is the stopping point? - A reduction in R^2_{adj}
or an increase in AIC

- Perform the additional sum of squares F tests

Rule: Take the step in forward selection
if F is significant
Take the step in backward elimination
if F is not significant

All possible subsets method

Fit every possible model & pick the one with the best R^2_{adj}, AIC, or C_p

Principal Components

- Create k new predictors, each of which is a linear combination of the original predictors, but the new predictors are completely uncorrelated with each other

Original model: $Y = X\beta + \varepsilon$

Compute $X'X$ & find its spectral decomposition

$$X'X = P \Lambda P'$$

(P: eigenvectors as columns; Λ: eigenvalues on diagonal)

Let $Z = XP$ so $X = ZP^{-1}$
$= ZP'$

Let $\gamma = P'\beta$ so $\beta = P\gamma$

$$Y = X\beta + \varepsilon = ZP'P\gamma + \varepsilon = Z\gamma + \varepsilon$$

Note: $Z'Z = (XP)'XP$
$= P'X'XP$
$= P'(P\Lambda P')P$
$= \Lambda$

Method: (1) Find the principal components

(2) Eliminate any that correspond to 0 eigenvalues

(3) Use any stepwise method or all subsets to find the best subset of these, as predictors

(4) Translate your final model back to the original predictors

Example: Predictors X_1, X_2, X_3 but they have a multicollinearity problem

Create $Z_1 = a_1 X_1 + a_2 X_2 + a_3 X_3$

$Z_2 = b_1 X_1 + b_2 X_2 + b_3 X_3$

$Z_3 = c_1 X_1 + c_2 X_2 + c_3 X_3$

Suppose that the procedures keeps Z_1 & Z_2 in the model

$\hat{Y} = \hat{\gamma}_0 + \hat{\gamma}_1 Z_1 + \hat{\gamma}_2 Z_2$

$$\hat{Y}_i = \hat{\gamma}_0 + (\hat{\gamma}_1 a_1 + \hat{\gamma}_2 b_1) X_1 + (\hat{\gamma}_1 a_2 + \hat{\gamma}_2 b_2) X_2 + (\hat{\gamma}_1 a_3 + \hat{\gamma}_2 b_3) X_3$$

Grand scheme:

- Plot Y vs. each single predictor to decide what power(s) of that predictor should go into the model
- Decide which cross-products would be appropriate as predictors

- Examine possible transformations of Y
- Model selection (all possible subsets or stepwise)
- residual analysis (error plots, leverage & influence)